Anchoring of semiotic symbols

Vogt, P.

Published in:
Robotics and Autonomous Systems

Publication date:
2003

Citation for published version (APA):
Vogt, P. (2003). Anchoring of semiotic symbols. *Robotics and Autonomous Systems*, *43*(1), 109-120.

Anchoring of semiotic symbols

Paul Vogt

Institute for Knowledge and Agent Technology, Universiteit Maastricht, P. O. Box 616, 6200 MD Maastricht, The Netherlands

Abstract

This paper presents arguments for approaching the anchoring problem using *semiotic symbols*. Semiotic symbols are defined by a triadic relation between forms, meanings and referents, thus having an implicit relation to the real world. Anchors are formed between these three elements rather than between 'traditional' symbols and sensory images. This allows an optimization between the form (i.e. the 'traditional' symbol) and the referent. A robotic experiment based on adaptive language games illustrates how the anchoring of semiotic symbols can be achieved in a bottom-up fashion. The paper concludes that applying semiotic symbols is a potentially valuable approach toward anchoring.

Key words: Anchoring problem; symbol grounding problem; physical grounding; adaptive language games; semiotics

1 Introduction

The symbol grounding problem that deals with the question how symbols can be used meaningfully [8] is one of the hardest problems in AI and robotics. As many robotic applications use symbols for reasoning, problem solving and communication, solutions for this problem are extremely important for robotics

[1] Email: p.vogt@cs.unimaas.nl
The research has been carried out at the AI Lab of the Vrije Universiteit Brussel. I would like to thank Luc Steels very much for enriching me with his expert advice and inspiration. I am also indebted to all other colleagues of the AI Lab for their support in one way or another. Ida Sprinkuizen-Kuyper, Rens Kortmann and three anonymous reviewers are thanked for their critical comments that helped improve this paper.

research and development. But symbol grounding is also an important problem in studying foundations of cognition such as the evolution of language, as human language is primarily symbolic [7].

Recently a formalized solution for the technical aspect of the symbol grounding problem has been proposed under the name of *anchoring* [5]. Anchoring concentrates on constructing and maintaining a relation between a symbol and a sensory image that is acquired from observing a physical object. Symbol grounding is, in addition to anchoring, also concerned with 'anchoring' abstractions and, more fundamentally, with philosophical issues relating to the meaning of symbols.

Many attempts to tackle the anchoring problem start with the design of predefined symbol systems that have predefined anchors to relate symbols with visual percepts [5,13]. Recently, an increasing number of attempts have been made to approach the anchoring problem from the bottom-up in which robots develop their symbolic representations during their evolution - be it phylogenetic and/or ontogenetic. These attempts often relate to the development of symbolic communication [2,12,16,22,24].

The common approach to tackle the anchoring problem focuses on the development – hand-coded or learned – of anchors between symbols and sensory images [5]. This is a difficult problem since the robots have to deal with the object constancy problem: When viewing an object from different locations, the sensory images relating to this object differ enormously because the size of the projection may differ or because the object may be obscured. Humans are well capable of dealing with object constancy, but it is unclear how this works. One approach to tackle the problem of object constancy would be to develop anchors between symbols and the real world object, rather than between symbols and sensory images.

This paper proposes that the anchoring problem can be solved in terms of *semiotic symbols*, which have implicit anchors in the real world [22]. An experiment based on Steels' language game model [14] illustrates how anchors in these semiotic symbols may be constructed from the bottom-up through the use of language. In addition, it is discussed how the presented language game model may explain the cognitive phenomenon of family resemblance [23].

The paper is organized as follows: The next section presents the notion of semiotic symbols and discusses some of the requirements for anchoring these. The experimental setup is presented in section 3. Section 4 presents the experimental results. Discussions of the issues raised in the paper are presented in section 5. Conclusions are given in section 6.

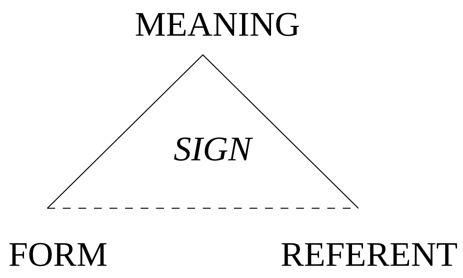

Fig. 1. The semiotic triangle illustrates the relations between referent, form and meaning that constitute a sign. Each line is an anchor, but the dotted line indicates that the relation between a form and a referent need not be a physical anchor, which must be established between referent and meaning and between meaning and form.

2 The anchors of semiotic symbols

In this section, I will define the notion of semiotic symbols as opposed to the definition of symbols that is commonly used in AI. As I will argue below, semiotic symbols have implicit anchors between some internal structures and reality. Finally, I will discuss under what conditions semiotic symbols may emerge.

The definition of semiotic symbols is adopted from Peirce [11], who defined a semiotic symbol in terms of a sign, which in semiotics is a relation between a *referent*, *meaning* and *form.*[2] These three elements can be described as follows:

Form A form (or *word*) is the shape of the sign, which is not necessarily material.
Meaning The meaning is the sense that is made of the sign.
Referent A referent is the object that stands for the sign, which may include abstractions, actions or other signs.

The relation between the referent, form and meaning is often illustrated with the semiotic triangle [10] as shown in Fig. 1. According to Peirce, a sign becomes a *(semiotic) symbol* when its form, in relation to its meaning is arbitrary or conventionalized so that the relationship has to be learned; otherwise the sign is either an *icon* or an *index.*

A semiotic symbol becomes meaningful when it is constructed and used functionally by an agent, which is conform Wittgenstein [23]. As such the meaning

[2] Peirce called this a symbol rather than a semiotic symbol. I call it a semiotic symbol to distinguish it from the – in AI and some other disciplines of cognitive science – commonly used definition of a symbol, which is similar to the form of the semiotic symbol. In addition, Peirce used the terms *representamen*, *interpretant* and *object* where I use the terms form, meaning and referent.

arises from the interaction of an agent that uses a form with the referent. Elsewhere I have argued that the symbol grounding problem as presented by Harnad is no longer relevant when we adopt semiotic symbols, because these are *per definition* grounded as their meanings have intrinsic relations with their referents [22]. This, however, does not solve the symbol grounding problem, but translates it into another – more technical – problem, which I have coined the *physical symbol grounding problem.* [3]

The physical symbol grounding problem is related to the anchoring problem in that it aims at constructing and maintaining anchors between symbols – i.e. the *forms* in semiotic symbols – and reality. Coradeschi and Saffiotti's description of anchoring, however, focuses on anchors between forms and sensory data [5]. As the sensory data is acquired from a robot's interaction with its its environment, the forms relate to the real world. The anchors, however, are not necessarily constructed to maintain a relation with the real world entity, but rather with the sensory image of this entity. The physical symbol grounding problem, on the other hand, does focus on constructing and maintaining a relation with the real world by constructing anchors between forms and real world entities, mediated by anchors between forms and meanings and between meanings and referents. In addition, where anchoring relates forms to sensory images (and thus to the sensing of physical objects), the physical symbol grounding problem is not restricted to constructing semiotic symbols about physical objects, but also include abstractions, movements and even other semiotic symbols.

The development of semiotic symbols depends on how an agent interacts with its environment. When the semiotic symbols are used in language, the way the meaning is constructed depends on how it is used [23]. However, the meaning of semiotic symbols also must have a part that can be memorized, which can be represented in terms of prototypical categories. When mediating on the meaning of a semiotic symbol, agents must confer to a similar meaning, hence they must try to find a common way to name the meaning. It is not unlikely that this requires for the agents to construct similar representations of the meanings they use. In addition, the construction of semiotic symbols should be adaptive, because it may be impossible to design 'static' anchors that apply to the dynamic interactions of a robot with its environment [9]. An adaptive approach to construct semiotic symbols allows robots to create new anchors when none exist or when existing ones are insufficient. As a result, I assume that a semiotic symbol can have multiple meanings (or prototypes) to stand for a referent in relation to a form. These different meanings of a semiotic

[3] This problem is coined the physical symbol grounding problem to indicate that semiotic symbols provide a way to approach symbol grounding with the physical grounding hypothesis [4] as the semiotic symbols themselves form a coupling between the environment and an agent's behavior and thus are physically grounded.

symbol will then be used to interpret a referent on different occasions. To achieve such a development of semiotic symbols in communication, I assume that the meanings co-develop with linguistic forms [3] by means of cultural interactions between agents and their environment [18].

The anchors between meanings and referents arise from the physical interactions between an agent and its environment. The meanings are anchored to linguistic forms through the production and interpretation of expressions. These physical anchors between referents, meanings and forms provide an implicit non-physical anchor between the forms and referents through their use in language (Fig. 1). The way these anchors are formed is influenced by the agents' interactions with their environment and individual adaptations as a self-organizing process [14].

For robots that develop semiotic symbols from the bottom-up, the above requires that robots are capable of interacting with their environment, including each other. Furthermore, they have to construct and memorize categorizations that provide anchors between the referents and the categories such that these can be used appropriately in language. To use these in language they also have to construct anchors between the categories and linguistic forms adaptively. How this can be modeled is explained in the next section.

3 Adaptive language games

To illustrate how a set of anchored symbols can be developed from the bottom-up, an experiment is presented in which two mobile LEGO robots bootstrapped a symbolic communication system. To achieve this, the robots engaged in a series of *adaptive language games* [14,17] in which they tried to communicate the form that stands for an object and adapt their internal structures in order to improve their performance on later occasions. Various types of language games have been implemented such as *observational games*, *guessing games* and *selfish games*, which differ from each other in the type of learning mechanism the robots use and in what non-verbal input they use to determine the reference of an utterance [19,20]. For the experiment of this paper, the robots played a series of *guessing games*. Below follows a technical description of the experimental setup.

3.1 The environment

In the experiment two mobile LEGO robots were used that were equipped with light sensors, bumpers, active infrared, two motors, a radio module and

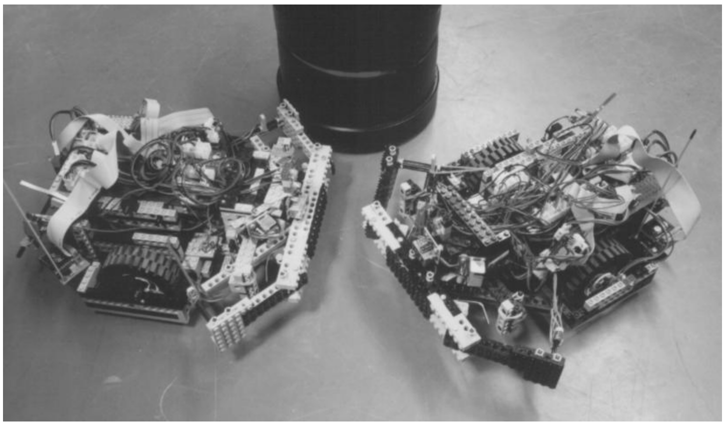

Fig. 2. The LEGO robots and a light source as used in the experiment.

a sensorimotor board, see Fig. 2. The light sensors were used to detect the objects in the robots' environment. The other sensors and the motors were used to process the physical behaviors of the robots.

The robots were situated in a small environment ($2.5 \times 2.5m^2$) in which four light sources were placed at different heights. The light sources acted as the objects that the robots tried to name. The four light sensors of the robots were mounted at the same height as the different light sources. Each sensor outputs its readings on a *sensory channel*. A sensory channel is said to *correspond* with a particular light source if the sensor has the same height as this light source.

The goal of the experiment was that the robots developed a lexicon with which they could successfully name the different light sources.

3.2 Sensing, segmentation and feature extraction

Through the interactions of the robots with their environment, they obtain raw sensory data. In order to reduce the redundant information from this high dimensional data, the robots transfer this data into low dimensional *feature vectors*. The process of acquiring feature vectors was done by *sensing*, *segmentation* and *feature extraction*. Each subsequent step reduced the amount of sensory data as if it were a sieve.

3.2.1 Sensing

A guessing game started when both robots were standing close to each other with their backs 'facing' each other.[4] During the sensing phase, the robots

[4] In the original implementation, the robots aligned themselves autonomously [17], but to speed up the experiments, the robots were placed by hand for this experiment.

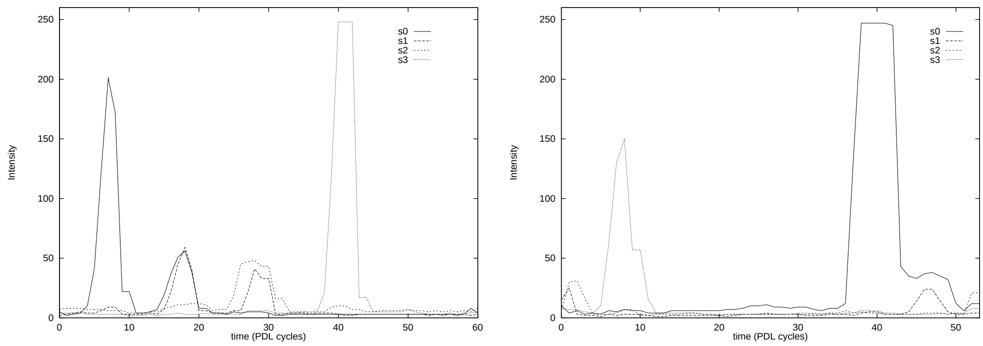

Fig. 3. The sensing of robot A (left) and robot B (right) during a language game. The plots show the spatial view of the robots' environment. It is acquired during 360^o of their rotation. The y-axis shows the intensity of the sensors, while the x-axis determines the time (or angle) of the sensing in PDL units. A PDL unit takes about $\frac{1}{40}$second, hence the total time of these sensing events took about $1.5s$ for robot A and $1.3s$ for robot B.

rotated one by one 720^o to obtain a spatial view of their environment. A spatial view contained the raw sensory data from the middle 360^o, which can be written in the form of a matrix[5]

$$X = \begin{bmatrix} x_{1,1} & \dots & x_{1,q} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,q} \end{bmatrix} \tag{1}$$

where each row represents the sensory data of the n sensory channels (4 in the experiment) and the detection of q measurements are given in the columns.[6] The sensory data was sent to a stand alone PC where all further processing took place off-line.

Figure 3 shows the sensing of the two robots during a guessing game. The left figure shows that robot A clearly detected the four light sources; there appears a 'winning' peak for every light sensor s_i that corresponds to one light source. The right figure shows that robot B did not sense all four light sources clearly and thus acquired a different view than robot A. This happened because both robots were not located at the same position.

[5] The robots rotated twice instead of once to ensure they rotated at a constant speed when the actual sensing started. This is done because the onset and offset of the movement induced a warped view, which in turn induced much noise for the segmentation.

[6] Note that although the robots have more than four sensors only the four light sensors are used to construct anchors.

3.2.2 Segmentation

The segmentation phase extracted connecting regions where the sensory data exceeded a threshold that represented the upper noise level of that sensor. These regions were supposed to be induced by the sensing of a light source. To accomplish this segmentation, the raw sensory input X was thresholded for noise resulting in $X' = \text{matrix}(x'_{i,j})$ according to

$$x'_{i,j} = H(x_{i,j} - \Theta_i) \tag{2}$$

where

$$H(x) = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases} \tag{3}$$

and Θ_i represents the upper noise level of light sensor i, which was acquired empirically for each sensor.

Given the preprocessed sensory data X', a segment S_k can be defined as the *largest* matrix

$$S_k = \begin{bmatrix} x'_{1,r} & \dots & x'_{1,m} \\ \vdots & \ddots & \vdots \\ x'_{n,r} & \dots & x'_{n,m} \end{bmatrix} \tag{4}$$

where in each column j there is at least one element for which $x'_{i,j} > 0$ for $i = 1, \dots, n$ and $j = r, \dots, m$; and where $1 \leq r < m < q$. Note that the inequality $r < m$ implies that the segments have to contain at least two measurements to filter out further noise. When a segment was detected at the start of the view and another was detected at the end, both segments were concatenated.

Ideally, the segmentation resulted in a set that contained a segment for each light source. This set constituted what is called the *context* of the guessing game, i.e. $\text{Cxt} = \{S_1, \dots, S_N\}$, where N is the number of segments that were sensed. Each robot participating in the guessing game acquired its own context which could differ from another.

3.2.3 Feature extraction

The feature extraction results in a feature vector $\mathbf{f} = (f_1 \ldots, f_n)$, where $f_i = \varphi(S_k)$ is a function that normalizes the maximum intensity of a sensory channel i to the overall maximum intensity within a segment S_k. I.e. the maximum value in row i of the matrix S_k is normalized to the maximum value of the entire matrix. Mathematically the function $\varphi(S_k)$ is given by:

$$\varphi(S_k) = \frac{\max_{j\in[r,m]}(x'_{i,j})}{\max_{S_k}(x'_{p,q})} \tag{5}$$

This way the function extracts the invariant property that the feature of the sensory channel with the overall highest intensity inside a segment has a value of 1, whereas all other features have a value ≤ 1. Or, in other words, the feature with value 1 corresponds to the light source the feature vector refers to. The space that spans all possible feature vectors $\mathbf{f}$ is called the n dimensional feature space $\mathcal{F} = [0,1]^n$, or *feature space* for short.

3.3 Discrimination game

Each robot played a *discrimination game* [15] to form a memorized representation of the meaning – or meaning for short – for each (potential) *topic*. A topic is a segment from the acquired context as described by its feature vector. The speaker selected its topic randomly from the context and this topic became the subject of communication. As the hearer of a guessing game tried to guess what the speaker's utterance referred to, it had to consider all segments in its context as a *potential topic*. A discrimination game was successful when it resulted in one or more categories that distinguished the topic from all other segments in the context. When the robot failed to find such a category, the discrimination game failed and the robot expanded its ontology in which the categories were stored. The discrimination game is a sequence of three processes: *categorization*, *discrimination* and *adaptation*.

3.3.1 Categorization

A category $c = \langle \mathbf{c}, \nu, \rho, \kappa \rangle$ was defined as a region in the feature space $\mathcal{F}$ and it was represented by scores ν, ρ and κ and a prototype $\mathbf{c} = (y_1, \ldots, y_n)$, where y_i were the coordinates of the prototype in each of the n dimensions of $\mathcal{F}$. The category was the region in $\mathcal{F}$ in which the points had the nearest distance to $\mathbf{c}$. Each feature vector in the context was categorized using the *1-nearest* *neighbor algorithm* [6]. So, feature vector $\mathbf{f}$ was categorized with that category c for which the prototype $\mathbf{c}$ had the smallest Euclidean distance $||\mathbf{f} - \mathbf{c}||$.

In order to allow generalization and specialization of the categories, different versions of the feature space $\mathcal{F}_\lambda$ were available to a robot. In each space a different resolution was obtained by allowing each dimension of $\mathcal{F}_\lambda$ to be exploited up to 3^λ times, where $\lambda = 0, \ldots, \lambda_{\text{max}}$. How this was done will be explained in section 3.3.3.

The use of different feature spaces allowed the robots to categorize a segment in different ways. The categorization of segment S_k resulted in a set of categories $C_k = \{c_0, \ldots, c_m\}$, where $m \leq \lambda_{\text{max}}$.

3.3.2 Discrimination

Suppose that a robot wants to find distinctive categories for (potential) topic S_t, then a distinctive category set DC can be defined as follows:

$$DC = \{c_i \in C_t \mid \forall (S_k \in \text{Cxt} \backslash \{S_t\}) : c_i \notin C_k\} \tag{6}$$

Or in words: the distinctive category set DC consists of all categories c_i of the topic S_t that are not a category of any other segment S_k in the context Cxt.

3.3.3 Adaptation

If $DC = \emptyset$, the discrimination game fails and the robot should adapt its ontology by constructing new categories. Suppose that the robot tried to categorize feature vector $\mathbf{f} = (f_1, \ldots, f_n)$, then new categories were created as follows:

(1) Select an arbitrary feature $f_i > 0$.
(2) Select a feature space $\mathcal{F}_\lambda$ that has not been exploited 3^λ times in dimension i for λ as low as possible. If no such space can be found, the adaptation is stopped.
(3) Create new prototypes $\mathbf{c_j} = (y_1, \ldots, y_n)$, where $y_i = f_i$ and the other y_r are made by combining the features from all existing prototypes in $\mathcal{F}_\lambda$.
(4) Add the new prototypical categories $c_j = \langle \mathbf{c}_j, \nu_j, \rho_j, \kappa_j \rangle$ to the feature space $\mathcal{F}_\lambda$, with $\nu = \rho = 0.01$ and $\kappa = 1 - \frac{\lambda}{\lambda_{\text{max}}}$.

The three scores ν, ρ and κ together constitute the meaning score $\mu = \frac{1}{3}(\nu + \rho + \kappa)$, which was used in the naming phase of the guessing games. Although the influence of this score was small, it helped to select a form-meaning association in case of an impasse. Where κ was kept constant, ν and ρ were increased when the category was distinctive (ν) and when it was used successfully in the naming phase (ρ); they were lowered otherwise. Exact details of these updates can be found in [20].

If the distinctive category set $DC \neq \emptyset$, the discrimination game was a success and the DC was forwarded to the naming phase of the guessing game. If a category c was used successfully in the guessing game, the prototype $\mathbf{c}$ of this category was moved toward the feature vector $\mathbf{f}$ of the topic:

$$\mathbf{c} := \mathbf{c} + \epsilon \cdot (\mathbf{f} - \mathbf{c}) \tag{7}$$

where $\epsilon = 0.1$ is a constant step size with which the prototype moved toward $\mathbf{f}$. This way the prototypes became more representative samples of the feature vectors it categorized.

The discrimination game as implemented here differs from the implementation of Steels [15] mainly in the representation and construction of categories. Steels used binary trees to split up the sensory (or feature) channels rather than using prototypes. The reason for using prototypes is that the world as sensed by a robot is not binary and splitting up categories in binary trees seems therefore inappropriate. In addition, Steels allowed categories to be formed in only one dimension or in any combination of the different feature dimensions; while in this implementation the categories were always n dimensional.

It is important to realize that all processing up to this point was carried out by each robot individually. This way, the ontologies, contexts and distinctive category sets differed from robot to robot.

3.4 Production

After both robots obtained distinctive categories of the (potential) topic(s), the speaker tried to communicate its topic based on its lexicon. The lexicon L was defined as a set of form-meaning associations: $L = \{\mathrm{FM}_i\}$, where $\mathrm{FM}_i = \langle F_i, M_i, \sigma_i \rangle$ was a lexical entry. Word-form F_i was made from an arbitrary combination of consonants and vowels taken from the alphabet, meaning M_i was represented by some category, and association score $\sigma_i \in \langle 0, 1 \rangle$ was a real number that indicated the effectiveness of the lexical entry based on past interactions. Each form could be associated with multiple meanings, and each meaning could have associations with more than one form.

The speaker of the guessing game ordered the distinctive category set DC based on the meaning score μ. It selected the distinctive category with the highest meaning score and searched its lexicon for form-meaning associations of which the meaning matched this distinctive category. If it failed to find such an element, the speaker first considered the next best distinctive category from the ordered DC. If all distinctive categories were explored and still no entry was found, the speaker could invent a new form as will be explained in section 3.7.

If there were one or more lexical entries that fulfilled the above condition, the speaker selected the entry that has the highest association score σ. The form that was thus produced was uttered to the hearer. In the on-board implementation this was done using radio communication, off-line the utterance was a shared variable.

3.5 Interpretation

On receipt of the utterance, the hearer searched its lexicon for entries for which the form matched the utterance *and* the meaning matched one of the distinctive categories of the potential topics. If it failed to find one, the lexicon had to be expanded, as explained in section 3.7.

If the hearer found one or more entries, it selected the entry that had the highest score $\Sigma = \sigma + \alpha \cdot \mu$, where $\alpha = 0.1$ is a constant weight. The potential topic that was categorized by this meaning was selected by the hearer as *the* topic of the guessing game. I.e. this segment was what the hearer guessed to be the subject of communication.

3.6 Corrective feedback

The effect of the guessing games was evaluated by the corrective feedback. If the speaker had no lexical entry that matched a distinctive category, or if the hearer could not interpret the speaker's utterance because it did not have a proper lexical entry in the context of the game, then the guessing game was a failure. The guessing game was successful when both robots communicated about the same referent. So if the hearer interpreted the utterance and thus guessed the speaker's topic, the robots had to evaluate whether they communicated about the same referent.

In previous work there have been various attempts to implement the corrective feedback physically as a pointing behavior. All these attempts, however, failed. In order not to focus too long on this problem and to prove the principle, it was assumed for the time being that the robots could do this and the verification was simulated. Naturally this problem needs to be solved in the future.

The corrective feedback was simulated by comparing the feature vectors of the two robots relating to their topics. If the features with value 1 matched for both topics, this means that the topics corresponded to the same referent and the guessing game was considered successful. If the hearer selected an inconsistent topic during the interpretation, then there was a *mismatch in referent* and the guessing game failed.

3.7 Lexicon adaptation

Depending on the outcome of the game, the lexicon of the two robots was adapted. There were four possible outcomes/adaptations:

(1) *The speaker had no lexical entry*: In this case the speaker created a new form and associated this with the distinctive category it tried to name. This was done with a certain probability, which was kept constant during the experiment at $P_s = 0.1$.
(2) *The hearer had no lexical entry*: The hearer adopted the form uttered by the speaker and associated this with the distinctive categories of a randomly selected segment from its context.
(3) *There was a mismatch in referent*: Both robots adapted the association score σ of the used lexical entry by $\sigma := \eta \cdot \sigma$, where $\eta = 0.9$ is a constant learning parameter. In addition, the hearer adopted the utterance and associated it with the distinctive categories of a different randomly selected segment.
(4) *The game was a success*: Both robots reinforced the association score of the used entry by $\sigma := \eta \cdot \sigma + 1 - \eta$. In addition, they lowered competing entries (i.e. entries for which either the form or the meaning was the same as in the used entry) by $\sigma := \eta \cdot \sigma$. The latter update is called lateral inhibition.

The coupling of the naming phase with the discrimination game and the sensing part makes that the emerging lexicon is grounded in the real world. The robots successfully solve the physical symbol grounding problem in some situation when the guessing game is successful, because only in those case a semiotic triangle (Fig. 1) is constructed completely in a functional – and thus meaningful – sense.

4 Experimental results

An experiment was done for which the sensory data of the sensing phase during 1,000 guessing games was recorded. From this data set it was calculated that the a priori chance for successful communication was 23.5% when the robots

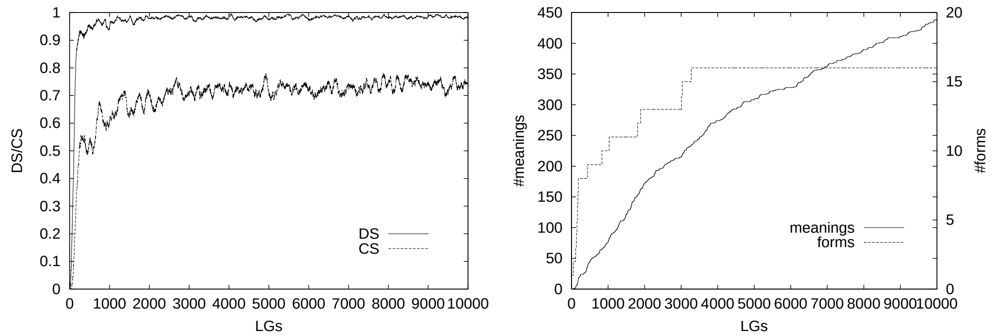

Fig. 4. (Left) The communicative success (CS) and discriminative success (DS) of the experiment. (Right) The evolution of the number of meanings and forms that were used successfully by the robots in one run of the experiment.

randomly chose a topic. Because the robots did not always detect all the light sources that were present, their context was not always coherent. This incoherence caused an upper limit to the success rate that could be reached, called the *potential understandability*, which was 79.5% on the average.

The 1,000 recorded situations were processed off-line on a PC in 10 runs of 10,000 guessing games. Figure 4 (left) shows the average communicative and discriminative success of the 10 runs. The communicative success measures the number of successful guessing games, averaged over the past 100 games. The discriminative success measures the number of successful discrimination games, also averaged over the past 100 guessing games. As the figure shows, the discriminative success reaches a value near 1 very fast. Hence, the robots were well capable of finding distinctive categories for the sensed light sources. The communicative success was somewhat lower. It increased toward a value slightly below 0.8 near the end. Since this is close to the potential understandability, the robots were capable to construct a shared lexicon within its limits.

Figure 4 (right) shows the number of different meanings and forms that were used at least once successfully in one run of the experiments. As the figure shows, the number of meanings used were much higher than the number of used forms. The robots used up to 450 meanings in relation to the four referents, while they only used 16 forms to name them. So, there are approximately 28× more meanings used than forms. Although the robots used about 450 meanings to distinctively categorize the four light sources, further analysis revealed they only used about 20 to 25 meanings frequently. In addition, only 6 or 7 forms were used regularly. So, the robots named each referent consistently with one or two forms.

The competition diagram of Fig. 5 (left) shows how the occurrence frequencies of the used forms to name one of the referents evolved during one run of the

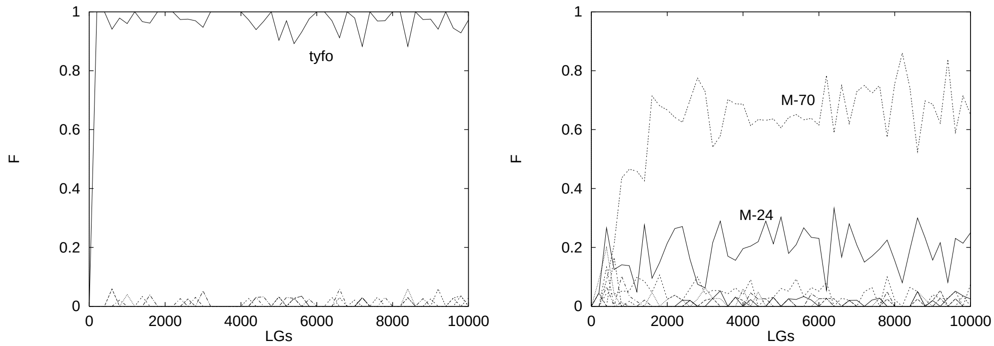

Fig. 5. The referent-form competition diagram (left) shows the competition between forms to name referent light source L1. The referent-meaning diagram (right) shows the competition between meanings to interpret light source L1. In both diagrams the y-axis shows the occurrence frequencies of successfully used forms or meanings over the past 200 games relative to the occurrence of the referent. The x-axis shows the number of games played.

experiment. As this figure makes clear, the most frequently used form "tyfo" clearly won the competition to name light source L1. At the bottom of the diagram, other forms reveal a weak competition. Similar competitions have been observed for the other referents [20,22]. Figure 5 (right) shows that the competition between meanings to categorize a referent is stronger, which would be expected given Fig. 4 (right). More experimental results can be found in [20,22].

5 Discussion

In this section, I will discuss why the notion of semiotic symbols is useful in relation to the anchoring problem. The discussion will be based on the observation that semiotic symbols can be constructed by optimizing the anchor between their forms and the objects they stand for; thus solving the object constancy problem. Furthermore, I will explain how the use of semiotic symbols can model the phenomenon of family resemblance.

In this paper the 'alternative' definition of symbols as semiotic symbols is adopted to provide the possibility to construct anchors between symbols (or forms as I call them) and the real world. But is there any advantage of using semiotic symbols over the traditional symbols in relation to the anchoring problem? In the original anchoring problem [5], anchors are sought between symbols and perceptual features, while the symbols' relations to the real world objects are somewhat brought to the background. The experiment of this paper revealed that it is the relation between the form and the real world object that is being optimized in terms of a one-to-one relationship. The relation between the form and the sensory data (or even the categories) does not reveal this optimization. I do not argue that the relationship between form and sensory data is unimportant, but I do want to argue that the relation between form and referent is the one we should care for.

Before explaining why the relation between form and referent is crucial, I will elaborate on the importance of the relation between sensory images and forms. The processes between sensing and feature extraction are extremely important because these transform the raw sensory data into more manageable feature vectors that additionally bear some invariant information concerning the referents. In addition, the intermediate representations of categories are important to allow the optimization between form and referent, because the discrimination games function – like the sensing, segmentation and feature extraction – as a sieve. This sieve enables the robots to bind the numerous variation of the sensing to more informative granules that are less numerous. These granules are, although still numerous, more manageable than the raw sensory data; thus allowing to close the coupling between referents, meanings and forms more easily.

The optimization between referent and form, however, is the most dominant process for the construction of consistent anchors between these two elements. To understand how this optimization works, it is important to realize that robots try to construct a lexicon that they can apply in different contexts. The lexicon is constructed through the interplay of adaptations under selective pressures and pragmatic language use. In the experiment, anchors were formed between referents and meanings, between meanings and forms; and between forms and referents. The results show that many anchors were used between referents and meanings, and between meanings and forms. However, when forms were used, they were well anchored to the referents they name. Failures in the discrimination game caused the emergence of so many meanings, because every time a discrimination game fails, a new category was added to the ontology. Many of them were associated with a form when they became distinctive in a later discrimination game. As associations were selected during a guessing game when their meanings fitted in the context – even if the scores were not high – a lot of these meanings were used successfully in the game.

The same context dependency causes the emergent tendency that the robots do not use so many forms, despite the variability of the acquired contexts during different games and between the robots. This can be understood by realizing that when one robot categorizes a referent differently in different guessing games, this does not necessarily mean that the other robot finds different distinctive categories. When the robot that uses the same distinctive category on different occasions, it will most likely use the same form to express this meaning too. This allows the other robot to use the form in association with the two different meanings successfully, as the game is context dependent.

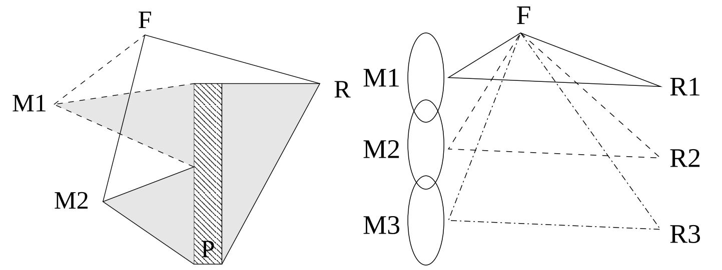

Fig. 6. Illustration of two semiotic relations between referent R, meaning M and form F. The left figure shows the continuum of possible sensing P of referent R are displayed as a rectangle. Some part of the rectangle may be interpreted by M1 and another by M2. When both meanings relate to the same form, this mechanism solves the problem of object constancy. The right figure shows how the model may explain family resemblance. The ovals should be interpreted as Venn-diagrams of the meanings M1 and M2.

When such situations occur frequently, this, in turn, allows the robots to use more meanings than forms. This emerging dynamics of the lexicon can be classified as semiotic dynamics and illustrates how conceptual development is, at least to some extent, dependent on language acquisition and language use; and vice versa. This is conform the – in a weaker version – revived Sapir-Whorf thesis [3]. A similar argument in favor of this weaker version of the Sapir-Whorf thesis was made in another study using language games [1]. In this study it was shown that agents developed a shared categorization of the color space when they used language, but a distinctive categorization when they developed categories without engaging in guessing games.

The optimization between referent and form solves, at least to some extent, the notion of object constancy: How can an object be recognized as being the same when different sensing of such an object can result in dramatically different sensory stimuli, for instance because it is partly obscured? Figure 6 (left) illustrates how the semiotic dynamics can explain the solution to the object constancy problem. In the experiment, the robots detected the light sources from different positions, resulting in different sensings – illustrated as the continuum of sensings P in Fig. 6 (left) – which may yield different meanings M1 and M2. Nevertheless, the system identifies the objects consistently, because the one-to-many relations between form and meaning converge at the level of form and referent.

The results of the experiment in this paper show that minimal autonomous robots can develop a shared set of semiotic symbols from the bottom-up by optmizing their anchors between forms and referents. However, one of the driving forces for this optimization – the corrective feedback – was simulated. This is a major shortcoming as the method used – inspecting each other's internal states – is unrealistic and may undermine the principle. Nevertheless, the assumption was adopted to test the principles of the underlying bootstrapping mechanisms and not to get stuck on solving this problem. A solution may come from applications were robots evaluate the corrective feedback using task-oriented behaviors, as was recently investigated in simulations [21]. In these simulations, the feedback came from the effect of the task that the agents had to perform using the evolved language.

The semiotic dynamics of the guessing games help to solve the object constancy problem, but it may also help to explain another interesting phenomenon observed in cognitive science, namely *family resemblance* [23]. Family resemblance is the observation that seemingly different things are called the same without being ambiguous, like the meaning of *games*. Where soccer and chess are typical games, a game like swinging is not typical. Swinging lies near the border of the 'conceptual space' of games – e.g. referent R1 in Fig. 6 (right). It has no direct resemblance with games like soccer and chess – e.g. R2 and R3 – but it has some resemblance with other games that in turn do have resemblance with soccer and chess. Such categorization process can be explained with the one-to-many relations between form and meaning. The word "games" is associated with different meanings for soccer, chess and swinging. The successful use of these meanings in different situated language games allows the system to emerge a family of resemblance. Optimization here should be made on the relation between a form and different referents. This optimization can be realized through the use of language.

Concluding, the above discussions provide many arguments in favor of using semiotic symbols over the traditional symbols with respect to anchoring. The most important argument is that in the construction of semiotic symbols, anchors between forms and reality are implicitly being optimized, rather than optimizing anchors between symbols and sensory images.

6 Conclusions

This paper illustrates how a small group of autonomous robots can develop a set of shared semiotic symbols in a bottom-up fashion by engaging in adaptive language games. The semiotic symbols the robots construct are defined by physical anchors between referents and meanings, and between meanings and forms, which yield a non-physical anchor between form and referent. The use of semiotic symbols allows a profitable optimization to find, track and (re)acquire anchors between forms and referents, rather than between forms and sensory images as proposed in the original description of the anchoring problem [5].

The experiments show how a consistent construction of semiotic symbols is positively influenced by their use in language. Through the use of language, the forms are shared externally to the robots. In addition, the robots share the reference of their communication through the received feedback. These external factors, together with the internal adaptations influence the way the robots organize their conceptual spaces. Thus their conceptual development is influenced to a large extent by their language use, hence providing an argument in favor of a weak interpretation of the Sapir-Whorf thesis as discussed in [3].

To further broad our understanding on the emergence of semiotic symbol systems in language use, additional research is required on the emergence of compositionality as this is one of the key aspects of human language use. Future research should concentrate on how compositional structures can be grounded in the sensorimotor flow through grammatical language use. In addition, more research is required to design robotic applications that are capable of verifying the effectiveness of their language use in order to provide corrective feedback autonomously. Although further research is required to improve and scale the model, adaptive language games provide a potentially valuable technology for a bottom-up approach toward anchoring semiotic symbols.

References

[1] T. Belpaeme. *Factors influencing the origins of colour categories.* PhD thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel, 2002.

[2] A. Billard. Imitation: a means to enhance learning of a synthetic proto-language in an autonomous robot. In K. Dautenhahn and C. L. Nehaniv, editors, *Imitation in Animals and Artifacts.* MIT Press, 2001.

[3] M. Bowerman and S. C. Levinson, editors. *Language acquisition and conceptual development.* Cambrige University Press, Cambridge, 2001.

[4] R. A. Brooks. Elephants don't play chess. *Robotics and Autonomous Systems*, 6:3–15, 1990.

[5] S. Coradeschi and A. Saffiotti. Anchoring symbols to sensor data: preliminary report. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000)*, pages 129–135, Austin, 2000.

[6] T.M. Cover and P.E. Hart. Nearest neighbour pattern classification. *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, 13:21–27, 1967.

[7] T. Deacon. *The Symbolic Species.* W. Norton and Co., New York, NY., 1997.

[8] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[9] D. Jung and A. Zelinsky. Grounded symbolic communication between heterogeneous cooperating robots. *Autonomous Robots*, 8:269–292, 2000.

[10] C. K. Ogden and I. A. Richards. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Routledge & Kegan Paul Ltd., London, 1923.

[11] C. S. Peirce. *Collected Papers*, volume I-VIII. Harvard University Press, Cambridge Ma., 1931. (The volumes were published from 1931 to 1958).

[12] D. K. Roy and A. P. Pentland. Learning words from sights and sounds: a computational model. *Cognitive Science*, 26:113–146, 2002.

[13] J. M. Siskind. Grounding the lexical semantics of verbs in visual pereption using force dynamics and event logic. *Journal of Artificial Intelligence Research*, 15:31–90, 2001.

[14] L. Steels. Emergent adaptive lexicons. In P. Maes, editor, *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulating Adaptive Behavior*, Cambridge Ma., 1996. The MIT Press.

[15] L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multi-Agent Systems*, Menlo Park Ca., 1996. AAAI Press.

[16] L. Steels and F. Kaplan. Aibo's first words. the social learning of language and meaning. *Evolution of Communication*, 4(1), 2000.

[17] L. Steels and P. Vogt. Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life*, Cambridge Ma. and London, 1997. MIT Press.

[18] M. Tomasello. *The cultural origins of human cognition*. Harvard University Press, 1999.

[19] P. Vogt. Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication*, 4(1):89–118, 2000.

[20] P. Vogt. *Lexicon Grounding on Mobile Robots*. PhD thesis, Vrije Universiteit Brussel, 2000.

[21] P. Vogt. Anchoring symbols to sensorimotor control. In *Proceedings of the 14th Belgian/Netherlands Artificial Intelligence Conference, BNAIC'02*, 2002. in press.

[22] P. Vogt. The physical symbol grounding problem. *Cognitive Systems Research Journal, special issue on Situated and Embodied Cognition*, 2002. in press.

[23] L. Wittgenstein. *Philosophical Investigations*. Basil Blackwell, Oxford, UK, 1958.

[24] H. Yanco and L. Stein. An adaptive communication protocol for cooperating mobile robots. In J-A. Meyer, H. L. Roitblat, and S. Wilson, editors, *From Animals to Animats 2. Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, pages 478–485, Cambridge Ma., 1993. The MIT Press.